

BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan sebuah studi mengenai hubungan antara satu variabel tak bebas (*regressand variable*) dengan satu atau lebih variabel bebas (*regressor variable*) (Gujarati & Porter, 2009). Hubungan yang diperoleh pada analisis regresi umumnya dinyatakan dalam bentuk sebuah persamaan matematik (Sudjana, 1996). Dengan demikian dapat disimpulkan bahwa analisis regresi adalah suatu teknik analisis data untuk mengetahui hubungan fungsional antara satu variabel tak bebas dengan satu atau lebih variabel bebas yang dinyatakan dalam sebuah persamaan matematik. Pada umumnya, variabel tak bebas dinotasikan dengan Y dan untuk satu atau lebih variabel bebas dinotasikan dengan X_1, X_2, \dots, X_k . Berdasarkan jumlah variabel bebasnya terdapat dua macam analisis regresi linear, yaitu regresi linear sederhana (*simple linear regression*) dan regresi linear berganda (*multiple linear regression*). Regresi linear sederhana merupakan teknik analisis data untuk mengetahui hubungan fungsional antara variabel tak bebas dengan satu variabel bebas, sedangkan regresi linear berganda merupakan teknik analisis data untuk mengetahui hubungan fungsional antara variabel tak bebas dengan dua atau lebih variabel bebas.

Penaksiran parameter pada model regresi linear sederhana dan regresi linear berganda menggunakan metode *Ordinary Least Square* (OLS). Metode OLS merupakan teknik penaksiran parameter dengan cara meminimalkan jumlah kuadrat kesalahan (*error*). Penaksir OLS merupakan penaksir tak bias linear yang paling baik atau disebut juga dengan *Best Linear Unbiased Estimator* (BLUE) dan juga penaksir OLS memiliki varians yang minimum dibandingkan dengan penaksir tak bias linear lainnya, sehingga parameter dari model regresi dapat ditaksir secara lebih akurat. Menurut (Gujarati & Porter, 2009), secara garis besar

dalam analisis regresi terdapat asumsi-asumsi yang harus dipenuhi agar penaksir analisis regresi dapat ditaksir dengan metode OLS, yaitu : (1) Normalitas, (2) Non-autokorelasi, (3) Homoskedastisitas, dan (4) Tidak terjadi multikolinearitas.

Menurut (Sembiring, 1995), model regresi umum untuk k variabel regressor adalah sebagai berikut :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

dimana Y menyatakan variabel tak bebas (*regressand variable*), X_i menyatakan variabel bebas (*regressor variable*) ke- i , β_i merupakan parameter koefisien regresi ke- i , dan ε merupakan kesalahan/*error*.

Multikolinearitas merupakan kondisi dimana terdapat hubungan linear yang nyata antar beberapa atau semua variabel bebas dari model regresi linear berganda (Gujarati & Porter, 2009). Terdapat dua jenis multikolinearitas, yaitu multikolinearitas sempurna dan multikolinearitas tidak sempurna. Multikolinearitas sempurna terjadi, apabila suatu variabel bebas bergantung sepenuhnya pada variabel yang lainnya (Sembiring, 1995). Apabila terjadi pelanggaran pada asumsi tidak terjadi multikolinearitas, maka taksiran parameter koefisien regresi dari variabel-variabel tersebut tidak dapat ditaksir dengan pasti. Dengan demikian, nilai taksiran parameter koefisien regresinya akan menjadi bias. Gejala kondisi multikolinearitas dapat dideteksi dengan menggunakan (1) Nilai *Variance Inflation Factor* (VIF) dan *Tolerance* (TOL), dan (2) *Eigenvalues* dan *Condition Index* (CI). Terdapat beberapa cara untuk mengatasi multikolinearitas yaitu informasi apriori, menggabungkan data *cross-sectional* dan data *time series*, mengeluarkan satu dari variabel yang berkolinear, transformasi variabel, penambahan data baru, mengurangi kolinearitas pada regresi polinomial, analisis komponen utama dan regresi ridge (Gujarati & Porter, 2009).

Regresi ridge adalah satu dari berbagai teknik analisis data untuk mengatasi permasalahan multikolinearitas pada datanya yang menghasilkan penaksir bias dari koefisien regresi (Kutner, Nachtsheim, Neter, & Li, 2005). Regresi ridge pada dasarnya merupakan metode OLS yang dimodifikasi dengan menambahkan

Ghaida Azzahra, 2020

**REGRESI RIDGE PARSIAL UNTUK DATA YANG MENGANDUNG MASALAH
MULTIKOLINEARITAS**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

sebuah konstanta bias c pada matriks korelasi $(X'X)$ dan variabel bebasnya ditransformasikan dengan prosedur *centering* dan *rescaling* (pemusatan dan penskalaan). Pemilihan konstanta bias c merupakan hal yang harus diperhatikan dalam regresi ridge. Konstanta c menggambarkan jumlah bias dalam penaksir $\hat{\beta}(c)$. Konstanta bias yang diharapkan adalah konstanta bias yang menghasilkan bias relatif kecil dan menghasilkan koefisien regresi yang relatif stabil. Suatu cara yang dapat digunakan dalam memilih konstanta bias c adalah dengan menggunakan nilai VIF dan *ridge trace* (jejak gulud).

Berikut beberapa penelitian terdahulu yang berkaitan dengan regresi ridge. T.L Wasilaine, M.W Talakua, dan Y.A Lesnussa melakukan penelitian pada tahun 2014 mengenai “Model Regresi Ridge untuk Mengatasi Model Regresi Linier Berganda yang Mengandung Multikolinearitas”, penelitian ini menghasilkan sebuah model regresi yang diperoleh dengan menambahkan konstanta bias c pada diagonal matriks $X^T X$. Kemudian Hasriani pada tahun 2014 melakukan penelitian yang berkaitan dengan “Perbandingan Regresi Ridge (Regresi Gulud) dan *Principal Component Analysis* (Analisis Komponen Utama) dalam Mengatasi Masalah Multikolinearitas”, hasil dari penelitian ini adalah metode regresi ridge lebih baik dari analisis komponen utama yang dilihat dari nilai MSE dan R^2 nya. Selanjutnya yaitu penelitian yang dilakukan oleh Nur Aeniatus Solekakh, Dwi Ispriyanti, dan Sudarno pada tahun 2015 dengan judul “Estimasi Parameter Regresi Ridge Menggunakan Iterasi Hoerl, Kennard, dan Baldwin (HKB) untuk Penanganan Multikolinearitas”, dari penelitian tersebut diperoleh sebuah model regresi dengan iterasi HKB sebanyak 64 kali.

Berdasarkan paparan di atas, pendekatan regresi ridge belum memberikan hasil yang benar-benar memuaskan dalam mengatasi masalah data yang mengandung multikolinearitas. Terdapat sejumlah masalah atau keterbatasan dalam regresi ridge, antara lain yaitu konstanta bias diterapkan untuk semua variabel bebas terlepas dari tinggi rendahnya tingkat multikolinearitas dan pemilihan konstanta bias c yang bersifat subjektif dan sembarang. Dengan adanya

Ghaida Azzahra, 2020

**REGRESI RIDGE PARSIAL UNTUK DATA YANG MENGANDUNG MASALAH
MULTIKOLINEARITAS**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

kelemahan dan keterbatasan dari regresi ridge, C.K. Chandrasekhar, H. Bagyalakshmi, M.R. Srinivasan, dan M. Gallo mengusulkan sebuah pendekatan yang dapat mengurangi keterbatasan pada regresi ridge, yaitu regresi ridge parsial. Regresi ridge parsial melibatkan penyesuaian secara selektif konstanta ridge terkait dengan tingginya variabel kolinear untuk mengontrol ketidakstabilan varians dari estimasi koefisien. Untuk memilih konstanta bias dari variabel-variabel kolinear, regresi ridge parsial menggunakan pendekatan *Singular Value Decomposition* (SVD). Berdasarkan uraian di atas, penulis tertarik untuk mengambil topik mengenai “Regresi Ridge Parsial Untuk Data Yang Mengandung Masalah Multikolinearitas”.

1.2 Rumusan Masalah

Berdasarkan uraian dari latar belakang maka rumusan masalah dari pelaksanaan penelitian ini adalah sebagai berikut :

1. Bagaimana prosedur penggunaan regresi ridge parsial dalam menangani masalah data multikolinearitas?
2. Bagaimana penerapan regresi ridge parsial pada kasus angka kematian bayi di Provinsi Jawa Timur tahun 2018?
3. Bagaimana perbandingan persamaan regresi ridge dengan persamaan regresi ridge parsial pada kasus angka kematian bayi di Provinsi Jawa Timur tahun 2018?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Uji multikolinearitas yang digunakan adalah dengan pendekatan nilai *Variance Inflation Factor* (VIF) dan *Tolerance* (TOL).
2. Kriteria pemilihan model terbaik menggunakan nilai *Mean Square Error* (MSE).

Ghaida Azzahra, 2020

**REGRESI RIDGE PARSIAL UNTUK DATA YANG MENGANDUNG MASALAH
MULTIKOLINEARITAS**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut :

1. Menjelaskan prosedur penggunaan regresi ridge parsial dalam menangani masalah data multikolinearitas pada regresi linear berganda.
2. Mengetahui penerapan regresi ridge parsial dalam menangani masalah data multikolinearitas.
3. Mengetahui hasil perbandingan persamaan regresi ridge dengan persamaan regresi ridge parsial.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut :

1. Memberikan dan menambah wawasan keilmuan kepada penulis dan pembaca khususnya dalam ilmu statistika mengenai regresi ridge parsial yang dapat diaplikasikan untuk menangani data yang mengalami masalah dengan tingkat multikolinearitas tinggi.
2. Memberikan informasi atau pengetahuan kepada pembaca mengenai penerapan regresi ridge parsial, khususnya dalam penelitian diterapkan pada kasus angka kematian bayi di Provinsi Jawa Timur tahun 2018, sehingga untuk instansi Dinas Kesehatan Provinsi Jawa Timur menjadi bahan informasi dan pertimbangan dalam menurunkan angka kematian bayi di Provinsi Jawa Timur.
3. Memberikan dan menambah wawasan keilmuan kepada penulis dan pembaca mengenai hasil perbandingan dari regresi ridge dan regresi ridge parsial untuk data yang mengandung masalah multikolinearitas tinggi.